When Anaphor Resolution Fails

William H. Levine, Alexandria E. Guzmán, and Celia M. Klin

State University of New York at Binghamton

Implicit in the study of anaphors has been the assumption that anaphoric inferences must be resolved for a text to be coherent. In six experiments we examined this assumption. Using recognition and reading time measures, we found that antecedents (e.g., "tart") were not accessed when an anaphor (e.g., "dessert") was read if the antecedent was well backgrounded and if a highly salient same-category distractor (e.g., "cake") intervened between the anaphor and its antecedent. We conclude that when an antecedent is difficult to retrieve, and when the failure to connect an anaphor to its antecedent does not create a coherence break, readers may simply read on rather than devoting additional time and attention to identifying the antecedent.© 2000 Academic Press

Key Words: inference; anaphor; reading; discourse.

Reading comprehension involves building a coherent mental representation of the information described by a text. A central aspect of this process involves making connections between currently read text and information from earlier in the passage. These connections are often signaled by anaphors, which are words or phrases that refer to some earlier-mentioned concept. The importance of anaphors in the comprehension of text is reflected by the extensive literature examining the process by which a reader makes a connection between an anaphor and its referent, or antecedent (e.g., Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983; Ehrlich & Rayner, 1983; Gernsbacher, 1989; Greene, McKoon, & Ratcliff, 1992; O’Brien, Duffy, & Myers, 1986; O’Brien, Raney, Albrecht, & Rayner, 1997; Sanford & Garrod, 1989).

The process of connecting an anaphor to its referent usually feels effortless to the reader. Despite this perception, finding the antecedent for an anaphor is an inference. That is, it is up to the reader to determine that the noun phrase “the burglar,” for example, refers to the same person who was introduced earlier in a story as “the criminal” (McKoon & Ratcliff, 1980). Just as with other types of inferences, anaphoric inferences require readers to incorporate information into their text representation that is not explicit in the text. However, despite the similarities, the study of anaphors has tended to differ in important ways from the study of other classes of inferences. Most notably, research examining causal, forward, and instrumental inferences, for example, has focused on determining if these inferences are drawn during reading, or more specifically, if they are drawn under an assortment of text conditions (e.g., Cotter, 1984; Klin, Guzmán, & Levine, 1999; Klin, Murray, Levine, & Guzmán, 1999; Klin & Myers, 1993; Klin, 1995). In contrast, in much of the research examining anaphoric inferences, the starting assumption has been that they are drawn and drawn reliably. Thus, rather than asking if anaphoric inferences are drawn, the central focus has been on examining the processes involved in drawing these inferences. And rather than examining the influence of various text conditions on the probability that anaphoric inferences are drawn, research has fo-
cused on the influence of various text conditions on the timing of the process (for a review, see Sanford & Garrod, 1989). The belief that anaphoric inferences are reliably drawn is based on the assumption that a reader’s goal is to build a coherent representation of the text, and if the antecedent for a pronoun, for example, is not identified (e.g., She went to the store), this will result in a comprehension failure (e.g., Who is she?). Thus, if readers are not initially successful in identifying an antecedent, they should persevere, resorting to alternate strategies for establishing coherence, such as drawing bridging inferences (e.g., Haviland & Clark, 1974; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). In support of this claim, there are a number of demonstrations that reading time increases when an anaphoric inference is difficult (e.g., O’Brien, Plewes, & Albrecht, 1990).

In contrast with these assumptions, however, we hypothesize that there may be conditions under which readers fail to resolve anaphors, even when an antecedent is available. Instead of viewing anaphor resolution as necessary, there may be conditions under which readers’ “standard of coherence” (van den Broek, Risden, & Husebye-Hartman, 1995) is met even when the referent of an anaphor is not identified. For example, in contrast with pronouns, which convey little information and are almost always used anaphorically, noun phrase anaphors contain more information (e.g., Fletcher, 1984) and can either be used anaphorically or to introduce a new concept (Fraurud, 1990; Poesio & Vieira, 1998). Consider the sample passage in Table 1. Although comprehension might be enhanced if readers make a connection between the categorical noun anaphor, “the dessert,” mentioned in the anaphor ending, and “the tart,” described earlier, the failure to make this connection may not be disruptive to the comprehension process. That is, readers may be content to know that some dessert exists without specifically identifying what it is. Thus, if no single entity from the discourse representation is sufficiently activated when

---

**TABLE 1**
Sample Text—Experiment 1

<table>
<thead>
<tr>
<th><strong>Introductions</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wanda was throwing a surprise party for her best friend John. John had just been promoted to Vice President of the company and some of his close friends wanted to congratulate him.</td>
</tr>
<tr>
<td><strong>Distractor version</strong></td>
</tr>
<tr>
<td>Wanda even made him a tart. She felt a little pressured because her daughter’s graduation was the next day and she needed to prepare for that occasion as well. She still had to get decorations and stop at the bakery for a cake. Her daughter loved their chocolate cakes. They were filled with a very rich cream and had gooey dark chocolate frosting on them. Wanda decided to have the cake decorated with white flowers and a white border.</td>
</tr>
<tr>
<td><strong>No Distractor version</strong></td>
</tr>
<tr>
<td>Meanwhile John was rushing around preparing for his daughter’s graduation the next day. He still had to stop at the store for decorations. Wanda was working hard preparing for the evening. She had bought flowers for the table and even made a tart.</td>
</tr>
</tbody>
</table>

| **Wrap-up** |
| Wanda hoped John’s party would be fun. The guests arrived right on time. As everyone sat down to eat, Wanda |
| Anaphor ending (Distractor and No Distractor versions) |
| said to save room for the dessert. |
| No Anaphor ending (Distractor version only) |
| congratulated John on the promotion. |

| **Recognition probe** |
| TART |

| **Comprehension question** |
| Was John recently fired? |
the anaphor is encountered, rather than engaging in additional processing, readers may simply read on.

Recent findings from O’Brien et al. (1997) provide initial support for this hypothesis. They used a rapid serial visual presentation paradigm, at a rate of 250 ms/word. After the anaphor was presented, participants were asked to perform a naming task on the adjective modifier of the target antecedent. When there was substantial distance between the anaphor and its referent, no facilitation was found. In the current experiments we go a step further and suggest that even when readers are allowed to read at their own pace they may not identify the antecedent for an anaphor. Further, we explore the influence of a variety of text factors on the resolution process. We hypothesize that readers should fail to resolve anaphors (1) when the antecedents are difficult to retrieve and (2) when this resolution failure does not create a break in the coherence of the passage. When both of these conditions are met, it should be inefficient for readers to invest attentional resources to retrieve the antecedent. In contrast, when antecedents are easily accessible, readers should resolve regardless of the need to do so. Similarly, if the failure to locate an antecedent for an anaphor creates a break in the coherence of the narrative, readers should be more likely to devote the resources needed to retrieve the antecedent.

In the current set of experiments we tested these hypotheses by creating passages in which antecedents should be difficult to retrieve from memory and in which the failure to resolve the anaphors should not create a break in the coherence of the text. In a series of six experiments, participants read passages, such as the one in Table 1, that either ended with a noun phrase anaphor (e.g., dessert) or had a control ending. An antecedent (e.g., tart) was always present in the passage. The primary question was if conditions exist under which readers do not resolve the anaphor. In addition to contributing to our understanding of anaphoric inferences, these experiments provided a test of the factors that influence the accessibility of information in long-term memory and of the factors that influence readers’ assessment of coherence.

We start by considering the accessibility of antecedents. Given a bottom-up theory of text comprehension, such as Kintsch’s (1988) Construction-Integration model or Myers and O’Brien’s (1998) resonance model, the accessibility of an antecedent should be influenced by factors such as the distance between the anaphor and the antecedent, the amount of elaboration of the antecedent, and the degree of match between the anaphor and its antecedent (Myers & O’Brien, 1998). This is based on the assumption that anaphors act like any other input to memory, automatically activating related information in memory. All related concepts in memory resonate in parallel (cf. Gillund & Shiffrin, 1984; Hintzman, 1986; Ratcliff, 1978) and the degree to which potential antecedents resonate should depend on their level of activation as well as the number of semantic and conceptual features that overlap with the anaphor.

The first text factor that we included was substantial referential distance between the anaphor and its antecedent. In the passages used in the current experiments, several sentences intervened between an anaphor and its antecedent. Moreover, the intervening material constituted a shift in topic. Thus, when the anaphor was encountered, the antecedent was always backgrounded, or out of focus. Given any model of memory, increasing the amount of time between encountering the antecedent and encountering the anaphor should decrease the accessibility of the antecedent (e.g., Clark & Sengul, 1979; Duffy & Rayner, 1990; Ehrlich & Rayner, 1983; O’Brien, 1987). Distance may decrease the activation of the antecedent either because its strength decays over time or because more recently mentioned concepts are likely to overlap with the contents of working memory (Myers & O’Brien, 1998).

The second text factor we explored was the presence of other potential antecedents, or distractors. As Greene et al. (1992) suggested, failure to locate a referent may occur because either no entity matches well or because several entities match about equally well. In the passages used in the current experiments, when a same-category distractor was included in the passage, two entities shared semantic features
with the anaphor. Thus, when the anaphor was read, its signal should have been divided between the referent and distractor, reducing the activation of the referent (Myers & O’Brien, 1998) by creating a fan effect (Anderson, 1976; Myers, O’Brien, Balota, & Toyofuku, 1984; Reder & Ross, 1983).

In support of this, Corbett (1984; see also Corbett & Chang, 1983; Mason, 1997) found that reading times were longer on an anaphoric phrase (e.g., frozen vegetables) when the passage contained both a referent (e.g., frozen asparagus) and a distractor from the same category (e.g., fresh corn) than when it only contained a referent. Corbett argued that both potential antecedents were initially accessed when the anaphoric phrase was read, and the correct referent was eventually identified by using the information provided by the prenominal adjective (i.e., frozen). Whereas the inclusion of a distractor has been shown to slow comprehension, it is not clear whether it will also eliminate resolution.

The third factor we examined was the salience of the distractor. This was manipulated in two ways. First, the amount of elaboration of the distractor was varied. A large amount of elaboration should serve to increase the probability that the distractor is encoded and produce a richer, more highly integrated memory representation. When the anaphor is encountered, activation should be drawn away from the antecedent. Support for this hypothesis was found in O’Brien et al. (1990); an elaborated antecedent was retrieved more quickly than an unelaborated antecedent even though the elaborated antecedent was more distant. In the current experiments an elaborated distractor may either cause readers to erroneously reinstate the distractor or to reinstate neither entity.

Salience was also manipulated by varying the typicality of the distractor. When the distractor is a more typical category exemplar (e.g., cake) than the referent (e.g., tart), it will share more semantic features with the category label (e.g., dessert) than the referent (Rosch, Mervis, Gray, Johnsen, & Boyes-Braem, 1976). This should cause the distractor to resonate more strongly because of its greater featural overlap with the anaphor, which, in turn, should make retrieval of the antecedent more difficult. Again, whereas the degree of match has not been shown to influence the probability of resolution, it has been shown to influence the time course of comprehension. Duffy and Rayner (1990; see also Garrod & Sanford, 1977) found that readers took longer to resolve categorical anaphors (e.g., bird) when the antecedent was an atypical member of the category (e.g., goose) than when it was a highly typical member (e.g., robin).

In addition to manipulating the accessibility of the antecedents, we also attempted to use anaphors that did not demand resolution. According to the bottom-up framework introduced earlier (Kintsch, 1988; McKoon, Gerrig, & Greene, 1996; Myers & O’Brien, 1998), if no entity is sufficiently activated when an anaphor is read, or if several entities match the anaphor about equally well, readers may continue processing by refocusing on the contents of working memory and sending new signals to memory until an antecedent is retrieved or by engaging in some sort of problem solving (Myers & O’Brien, 1998). Further, this continued processing should be most likely to occur when the failure to retrieve the antecedent makes it difficult to integrate the anaphor into the passage. In contrast, when a noun phrase can be easily integrated into the ongoing text representation without access to its referent, it may be more efficient to continue reading than to invest additional time and attention continuing to process the anaphor.

Without a full understanding of what makes a text coherent, it is difficult to define precisely the set of factors that should influence the likelihood that an anaphoric noun phrase can be comprehended without retrieving its antecedent. However, both the explicitness of the anaphor and the degree of contextual constraint should have an influence. In the current experiments, noun phrase anaphors (e.g., the dessert) were used rather than pronominal anaphors (e.g., she). The greater semantic content of noun phrases should allow them to more easily stand on their own; whereas written pronouns can only be used anaphorically, noun phrases can either be used anaphorically or to introduce a
new concept. And second, anaphors were used that were constrained by the passage context. For example, because “dessert” is part of our dinner party schema, it should be in “implicit focus” (Sanford & Garrod, 1981) when the passage is read (Garrod & Sanford, 1983; Walker & Yekovich, 1987). Therefore, even if there were no antecedent explicitly included in the passage, it could be used to refer to some non-specific dessert.

In summary, in the six experiments reported here we asked if there are any conditions under which noun anaphors fail to be resolved, even when an antecedent is available in the passage. Passages were constructed so that the antecedents should have been difficult to retrieve and so that the failure to resolve the anaphors should not have created a break in the coherence of the passages. In Experiment 1, after processing an anaphor (e.g., dessert), there was no evidence of resolution when the antecedent was back-grounded by a description of a highly typical, same-category distractor (e.g., cake). This was based on the failure to find facilitation on a recognition probe representing the antecedent concept (e.g., tart). In Experiments 2 and 3 we investigated the influence of referential distance and of the salience of the same-category distractor, respectively. Experiment 4 replicated the findings from Experiment 1 with reading times instead of recognition. The results of Experiment 5 ruled out the possibility that readers had actually resolved the anaphor by erroneously selecting the distractor as the referent, and in Experiment 6 an instructional manipulation was used to demonstrate that despite the resolution failures in Experiments 1 and 4, the antecedent was accessible and could be identified if readers devoted additional time.

**EXPERIMENT 1**

In Experiment 1, we asked if there are any conditions under which readers fail to resolve categorical anaphors when the antecedent is accessible and readers are free to read at their own pace. In contrast to the assumption that anaphoric inferences are reliably drawn, if the failure to resolve an anaphor does not constitute a coherence break, and the retrieval of the antecedent is difficult, readers may leave the anaphor unresolved.

Passages appeared in three versions (see Table 1). The first was a Distractor–Anaphor version, in which passages ended with a line that contained a categorical anaphor in a definite noun phrase (e.g., the dessert) that referred to an antecedent (e.g., tart) that had been mentioned earlier in the passage. To determine if there are conditions under which anaphors are not resolved, three text factors were manipulated in an attempt to decrease the accessibility of the antecedent. First, a same-category distractor was included in the passage. Second, the distractor was more salient than the antecedent. In the sample passage in Table 1, there is only one short sentence mentioning the tart, but there are four sentences related to the cake. In addition, the distractor was a more typical member of the category instantiated by the anaphor than the antecedent (Battig & Montague, 1969). Assuming that the distractor is activated when the anaphor is processed (e.g., Corbett, 1984), both the extensive elaboration and the high typicality should increase its level of activation and consequently decrease the level of activation of the antecedent. Third, the antecedent was thoroughly back-grounded when the anaphor was encountered. Not only was there a substantial amount of material intervening between the antecedent and the anaphor, but this material constituted a shift in topic. In the sample passage the intervening material shifted the topic of the narrative from the dinner party to the daughter’s graduation.

Immediately after the last line of the passage a probe word, the antecedent, was presented for recognition. If the anaphor is resolved, the antecedent should be highly active, leading to facilitation when it is presented as a test probe (e.g., Dell et al., 1983). In contrast, if readers fail to resolve the anaphor, responding to the antecedent probes should not be facilitated.

The second version, Distractor-No Anaphor, was identical to the Distractor-Anaphor version except that the last line did not contain an anaphor. Instead, it continued with the theme of the passage (“Wanda congratulated John on the promotion”) and in no way made reference to
the antecedent concept. This version served as a control. If readers resolve the anaphor in the Distractor-Anaphor version, recognition decisions should be faster and more accurate than in the Distractor-No Anaphor version. On the other hand, if they fail to resolve the anaphor, there should be no difference between the two versions.

Finally, in the third version, No Distractor-Anaphor, the last line contained the categorical anaphor, as it did in the Distractor-Anaphor version, but the passages were written so that the antecedent should be easily accessible. There was less intervening material between the anaphor and the antecedent and there was no same-category distractor. The lengthy elaboration about the cake, for example, was replaced with a shorter section of neutral material. Given the findings from several prior studies (e.g., Dell et al., 1983; O’Brien et al., 1997), the anaphor should be easily resolved in this version; the anaphor is unambiguous and the antecedent is only backgrounded by two to three lines of text. This condition was included as an additional baseline. If the anaphor is not resolved in the Distractor-Anaphor version, latencies and error rates should be the same as in the Distractor-No Anaphor version, leading to a null result. In contrast, latencies should be faster and error rates lower in the No Distractor-Anaphor version. This condition was also included to control for semantic priming; any priming from the anaphor (e.g., dessert) to the antecedent (e.g., tart) should be equal in the Distractor-Anaphor and No Distractor-Anaphor versions.

Method

Participants. Forty-eight undergraduates at the State University of New York at Binghamton participated as part of the requirement for an introductory psychology course. The data from four participants who made more than 25% probe or comprehension question errors were replaced. Therefore data analyses are based on 44 participants.

Materials. There were 18 experimental passages, with three versions of each: Distractor-Anaphor, Distractor-No Anaphor, and No Distractor-Anaphor. A sample passage can be found in Table 1. The Distractor-Anaphor and the Distractor-No Anaphor versions were identical until the last line: The antecedent was mentioned once early in the passage, followed by a topic shift where the distractor was introduced. The distractor was a member of the same category as the antecedent but was a more typical member than the antecedent (Battig & Montague, 1969). In addition, there were 4–7 lines of elaboration about the distractor. In the Distractor-Anaphor version, the last line of the passage included a categorical anaphor that referred to the antecedent. In contrast, in the Distractor-No Anaphor version the categorical anaphor was not included in the last line. Across versions the last lines were equated for length by number of characters. In the third version, No Distractor-Anaphor, the section describing the distractor was replaced by neutral filler information that was unrelated to the anaphor and was shorter. In addition, antecedents appeared an average of only 3 lines before the anaphor, in comparison to the two distractor versions in which the antecedents appeared an average of 10 lines before the anaphor.

All three versions of the passage were followed by a recognition probe, which was the antecedent. There were also 25 filler passages that varied in length. For each filler passage a recognition probe was generated. None of the filler passages ended with anaphors. For the 18 filler passages that were followed by probe words that required “no” responses, approximately 40% of the probe words were semantically related to the passages and the other 60% were unrelated words. Over the course of the experiment participants read a total of 43 passages, 25 of which required a yes probe response. A yes/no comprehension question was generated for each experimental and filler passage to encourage participants to read carefully. These questions never referred to the anaphor.

Design. For each participant, experimental passages were quasirandomly assigned to the Distractor-Anaphor, Distractor-No Anaphor, and No Distractor-Anaphor versions with two constraints: (a) Each participant saw one-third of the passages in each version and (b) each version of each passage was presented to one-
third of the participants. The order of presentation was the same for all participants. Filler passages were interspersed among the experimental passages.

Procedure. Participants were tested individually in an experimental session lasting approximately 45 min. The texts were presented one line at a time on a computer monitor. Participants were instructed to read for comprehension at their own pace. Each trial began with the presentation of the phrase READY FOR NEXT STORY at the center of the screen. Participants controlled the presentation of the text with a line advance key. Each key-press caused the current line to be erased and the next to be presented. Immediately after the final line of the passage had been presented, pressing the advance key caused the string “XXX” to be presented for 500 ms. Participants did not know they were at the end of the passage until this cue appeared. This was replaced by a recognition probe. Participants controlled the presentation of the text with a line advance key. Each key-press caused the current line to be erased and the next to be presented. Immediately after the final line of the passage had been presented, pressing the advance key caused the string “XXX” to be presented for 500 ms. Participants did not know they were at the end of the passage until this cue appeared. This was replaced by a recognition probe. Participants controlled the presentation of the text with a line advance key. Each key-press caused the current line to be erased and the next to be presented. Immediately after the final line of the passage had been presented, pressing the advance key caused the string “XXX” to be presented for 500 ms. Participants did not know they were at the end of the passage until this cue appeared. This was replaced by a recognition probe. Participants had been instructed to indicate, as quickly and as accurately as possible, whether the probe had been in the preceding text by pressing a YES or NO key on a response box. After responding, the probe word was erased from the screen and participants received feedback about their response time. On trials in which participants took longer than 1500 ms to respond, the phrase PLEASE TRY TO RESPOND TO PROBES MORE QUICKLY was presented for 1500 ms. Each probe was followed by the string “???” for 500 ms, which was followed by a Yes/No comprehension question. For trials in which the participants responded incorrectly, the word “Error” appeared in the center of the screen for 1500 ms.

Results and Discussion

In all experiments, $t_1$ and $F_1$ refer to analyses in which the error term is based on subject variability and $t_2$ and $F_2$ refer to analyses in which the error term is based on item variability. An $\alpha$ level of .05 was used across all experiments.

Probe recognition times. Mean correct recognition times are presented in Table 2. These values were computed after discarding outliers (Tukey, 1977), $^1$ 3.8% of the data. Most central

---

**Table 2**

<table>
<thead>
<tr>
<th>Version</th>
<th>Reading time</th>
<th>Probe response times</th>
<th>Probe error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1: Highly salient distractor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Distractor-Anaphor</td>
<td>1746</td>
<td>763</td>
<td>4.9%</td>
</tr>
<tr>
<td>Distractor-Anaphor</td>
<td>1779</td>
<td>804</td>
<td>14.0%</td>
</tr>
<tr>
<td>Distractor-No Anaphor</td>
<td>1775</td>
<td>804</td>
<td>15.5%</td>
</tr>
<tr>
<td><strong>Experiment 2: No distractor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaphor</td>
<td>1713</td>
<td>806</td>
<td>8.2%</td>
</tr>
<tr>
<td>No Anaphor</td>
<td>1736</td>
<td>874</td>
<td>8.2%</td>
</tr>
<tr>
<td>Difference</td>
<td>23</td>
<td>−68*</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Experiment 3: Less salient distractor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Distractor-Anaphor</td>
<td>1699</td>
<td>833</td>
<td>8.3%</td>
</tr>
<tr>
<td>Distractor-Anaphor</td>
<td>1720</td>
<td>824</td>
<td>9.2%</td>
</tr>
<tr>
<td>Distractor-No Anaphor</td>
<td>1658</td>
<td>876</td>
<td>11.0%</td>
</tr>
<tr>
<td><strong>Experiment 6: Highly salient distractor and instructions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaphor</td>
<td>2155</td>
<td>865</td>
<td>13.6%</td>
</tr>
<tr>
<td>No Anaphor</td>
<td>2027</td>
<td>907</td>
<td>14.4%</td>
</tr>
<tr>
<td>Difference</td>
<td>128*</td>
<td>−42*</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

$^*$ $p < .05$.  

---

$^1$ According to Tukey’s (1977) criterion, a score, $X$, is defined as an outlier if $X > H_U + (1.5) \times (H_U - H_L)$ or
to the hypotheses, probe recognition times for the Distractor-Anaphor and the Distractor-No Anaphor versions were identical. Thus, despite previous findings demonstrating that anaphoric inferences are reliably drawn, there was no evidence that readers resolved the anaphor in the Distractor-Anaphor version. After reading “save room for the dessert,” the concept “tart” was no more accessible in memory than after reading “congratulated John on the promotion,” despite the fact that it had been explicitly stated that Wanda was making a tart for dessert. The signal from the anaphor was clearly not strong enough to reactivate the intended antecedent. In addition, because the inference was not necessary for comprehension (i.e., readers did not need to know that the dessert was a tart), readers seem to have been willing to leave the referent unresolved rather than investing the resources needed to complete the resolution process. Consistent with McKoon and Ratcliff’s (1992) Minimalist hypothesis, it appears that readers did not draw an anaphoric inference because the antecedent was not easily available and the inference was not necessary for comprehension.

In contrast, recognition latencies were faster following the target line in the No Distractor-Anaphor version. That is, after reading “save room for the dessert,” participants were faster to respond to the probe “tart” when there was no distractor in the passage. A contrast of the No Distractor-Anaphor version with the average of the two distractor versions was significant by subjects, $t_1(43) = 3.52, SEM = 11.46, p = .001$, and by items, $t_2(17) = 2.16, SEM = 20.14, p = .04$. This provides an important comparison for the null effect for the two Distractor versions; clearly the task is sensitive to the activation level of the probed concept. Thus, we can conclude that if the anaphor had been resolved in the Distractor-Anaphor version, then it would have been reflected in faster recognition times.

**Error rates.** The pattern of error rates mirrors the pattern of recognition latencies. Error rates were almost identical in the Distractor-Anaphor and Distractor-No Anaphor versions, 14.0 and 15.5%, respectively, $ps > .5$. In contrast, in the No Distractor-Anaphor version where recognition latencies were faster, the error rate was also lower (4.9%). This was reliable when compared with the Distractor-Anaphor version, $t_1(17) = 2.83, SEM = .47, p = .012$, and $t_2(43) = 3.15, SEM = .174, p = .003$ and when compared with the Distractor-No Anaphor version, $t_1(17) = 3.24, SEM = .48, p = .005$, and $t_2(43) = 5.21, SEM = .12, p < .001$. Again, this provides a contrast for the null effect in the two distractor versions.

Although we have suggested that the faster latencies and lower error rates on the antecedent probe in the No Distractor-Anaphor version resulted from readers resolving the anaphor, a second possibility exists. The greater accessibility may have been the result of the antecedent appearing more recently in the passage than in the other two versions. The reduction in distance alone might have made the antecedent probe more accessible, even if readers had not resolved the anaphor. Although the data do not allow us to distinguish between the influence of anaphor resolution and the influence of distance in the No Distractor-Anaphor version, regardless of the reason for the greater accessibility we can conclude that the anaphor was not resolved in the Distractor-Anaphor version. This is based on the null effect in latency and error rates for the Distractor-Anaphor and the Distractor-No Anaphor versions. This lack of a difference demonstrates that the antecedent was no more active after the anaphor was encountered in the Distractor-Anaphor version than when no anaphor was encountered in the Distractor-No Anaphor version. Further, the shorter latencies and lower error rates in the No Distractor-Anaphor version provide evidence that the task was sensitive to the activation level of the probed concept, whatever the reason for the higher activation in this condition. And finally, in Experiment 4 we provide converging support for the claim that the anaphor was not resolved in the Distractor-Anaphor version.

---

$X < H_u + (1.5) \times (H_u - H_l)$. $H_u$ and $H_l$ refer to the scores that cut off the upper and lower 25% of the ranked scores, respectively. Further, note after data were discarded there were slight inequalities in numbers of observations per condition.
Reading times. Last line mean reading times are presented in Table 2. These were calculated after discarding outliers as defined by Tukey’s (1977) criterion; this eliminated 5.2% of the data. Reading times did not differ across versions, both $F$s < 1. Although we cannot draw strong conclusions from this null result, the lack of a slowdown on the anaphor line in the Distractor-Anaphor version suggests that readers did not attempt to retrieve the antecedent through a slow, effortful search of memory. Instead, we conclude that because no entity in memory resonated above a threshold, and the passage was coherent even without resolving the anaphor, readers had little incentive to invest additional resources.

EXPERIMENT 2

The results of Experiment 1 demonstrated that readers do not always resolve categorical anaphors, even when they are free to read at any rate and an antecedent is available. In Experiments 2 and 3 we examined some of the text factors that were responsible for the resolution failure. Three features were included in the passages in Experiment 1 in an attempt to reduce the accessibility of the antecedent: there was a substantial amount of backgrounding between the anaphor and its antecedent, a same-category distractor was included in the passage, and the distractor was highly salient; it was elaborated and it was a more typical category exemplar than the referent. In Experiments 2 and 3 we examined these factors individually to determine their contribution to the resolution failure.

In Experiment 2 we examined the role of referential distance. Simply having a lengthy episode intervening between the anaphor and its antecedent might have been sufficient to prevent resolution. Although noun phrase anaphors do not require their antecedents to be in focus, as do pronominal anaphors (Gordon, Grosz, & Gilliom, 1993), even a semantically rich definite noun phrase, such as “the dessert,” may not be sufficient to reactivate an entity (e.g., tart) that was mentioned only once and was backgrounded by approximately 100 words and an entire episode. In support of this, several studies have found a slowdown in the resolution process as the distance between the anaphor and its antecedent increased (e.g., Duffy & Rayner, 1990; Mason, 1997). Further, in a recent study (O’Brien et al., 1997) the referent for a categorical anaphor was not immediately activated when the anaphor (e.g., clothes) and its referent (e.g., baby clothes) were separated by approximately 60 words.

Passages appeared in two versions, Anaphor and No Anaphor. These were similar to the Distractor-Anaphor and Distractor-No Anaphor versions from Experiment 1 except that the distractor elaboration section was replaced by neutral filler material, equated for length with the distractor elaboration. This neutral material still backgrounded the antecedent by changing the focus of the passage, but it did not include a description of a same-category distractor. For example, in the sample passage (see Table 3) Wanda must still prepare for her daughter’s graduation but instead of needing to pick up a cake she must buy a necklace. The primary question is if the inclusion of backgrounding material that is completely unrelated to the anaphor will still reduce the accessibility of the antecedent and eliminate resolution.

Method

Participants. Forty-three undergraduates at the State University of New York at Binghamton participated for course credit or $5. The data from one participant who made more than 25% probe and comprehension question errors were replaced. Therefore data analyses were based on 42 participants.

Materials. There were 18 experimental passages, with two versions of each: Anaphor and No Anaphor. A sample passage can be found in Table 3. The passages were similar to the passages in the Distractor-Anaphor and Distractor-No Anaphor versions in Experiment 1 except that the distractor elaboration section was replaced by material that was unrelated to the anaphoric concept. This “neutral filler” was the same length as the distractor elaboration section from Experiment 1, was as similar as possible in its content, and constituted a change in topic. For example, in the sample passage, instead of describing Wanda’s quest for a chocolate cake
for her daughter’s graduation party, she is now in search of a pearl necklace.

Passage endings were identical to those in Experiment 1; in the Anaphor version, passages ended with the categorical anaphor (e.g., save room for the dessert) and in the No Anaphor version the last line continued with the theme of the passage and did not include the anaphor (e.g., congratulated John on the promotion). The last lines were equated for length by number of characters. Further, in both versions, the last line of the passage was followed by a recognition probe that was the antecedent in the Anaphor version (e.g., tart). The filler passages from Experiment 1 were used in this experiment.

Design and procedure. The design and procedure were identical to those in Experiment 1, except that there were two versions of the experimental passages instead of three.

Results and Discussion

Probe recognition times. The mean correct probe response times are presented in Table 2. These values were computed after discarding outliers (Tukey, 1977), which consisted of 3.3% of the data. In contrast to the pattern of results from Experiment 1, probe response times were 68 ms faster in the Anaphor version than in the No Anaphor version, 806 and 874 ms, respectively. This difference was significant by subjects, $t_1(41) = 6.33$, $SEM = 10.63$, $p < .001$, and by items, $t_2(17) = 5.17$, $SEM = 14.29$, $p < .001$. The faster recognition latencies in the Anaphor version show that readers resolved the anaphor, reinstating the backgrounded antecedent after reading the anaphor line (e.g., save room for the dessert).

Impressively, although the antecedent was mentioned only once and backgrounded by a shift in focus and approximately 100 words, the anaphor was resolved. The finding that distance was not sufficient to eliminate anaphor resolution is in contrast to the results of O’Brien et al. (1997), who found no}

\[\text{TABLE 3}\]

Sample Text—Experiment 2

<table>
<thead>
<tr>
<th>Introduction and antecedent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wanda was throwing a surprise party for her best friend John. John had just been promoted to Vice President of the company and some of his close friends wanted to congratulate him. Wanda even made him a tart.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neutral filler</th>
</tr>
</thead>
<tbody>
<tr>
<td>The day had been busy because Wanda’s daughter was graduating from high school tomorrow. She had needed to prepare for that as well. She had first stopped at the dry cleaner’s and then at the jeweler’s to pick up her daughter’s gift. Her daughter had asked for a pearl necklace for her graduation present. It had been a family tradition for quite some time to give the girls a pearl necklace when they graduated from high school.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wrap-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wanda hoped John’s party would be fun. The guests arrived right on time. As everyone sat down to eat, Wanda</td>
</tr>
<tr>
<td>Anaphor ending</td>
</tr>
<tr>
<td>said to save room for the dessert.</td>
</tr>
<tr>
<td>No Anaphor ending</td>
</tr>
<tr>
<td>congratulated John on the promotion.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>TART</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comprehension question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was John recently fired?</td>
</tr>
</tbody>
</table>

\[\text{WHEN RESOLUTION FAILS}\]
facilitation on an antecedent (e.g., baby clothes) when it was separated from the anaphor by approximately 60 words, substantially less than the distance between the anaphor and antecedent in the current experiment. Although it is difficult to pinpoint the critical differences between the two studies, in contrast with the current study, O’Brien et al. presented texts at a rate of 250 ms/word. Thus, resolution may have been delayed rather than eliminated. That is, if readers had been given additional time, they may have completed the resolution process.

The results of Experiment 2 are in contrast with the lack of a difference in recognition latencies for the Distractor-Anaphor and Distractor-No Anaphor versions in Experiment 1. Although this does not rule out the possibility that referential distance interacted with the same-category distractor to reduce the accessibility of the antecedent in Experiment 1, it does suggest that distance alone was not sufficient to eliminate anaphor resolution. The presence of the same-category distractor was therefore critical in reducing the accessibility of the antecedent. This hypothesis is investigated further in Experiment 3.

Error rates. Error rates are presented in Table 2. In contrast with the response latencies, error rates did not differ between versions. The fact that shorter response latencies in the Anaphor version were not accompanied by lower error rates is in contrast to the pattern of results in Experiment 1. This suggests that the lower error rates in the No Distractor-Anaphor version of Experiment 1 may have been the result of the decrease in the distance between the anaphor and the antecedent rather than the result of the anaphor being resolved.

Reading times. Last line reading times are presented in Table 2 and did not differ between versions. Although the last lines were not identical, they were equated for length. The lack of a slowdown in the Anaphor version therefore suggests that the anaphoric inferences did not involve a slow, strategic search of memory. Instead, because there was no distractor, the intended referent should have resonated more strongly than any other entity in memory and should have been quickly selected as the referent of the anaphor.

EXPERIMENT 3

The major conclusion from Experiment 2 was that the inclusion of a salient distractor was necessary for the resolution failure observed in Experiment 1; backgrounding alone was not sufficient. In the current experiment we ask if simply including any same-category distractor would have reduced the accessibility of the antecedent and eliminated anaphor resolution or if the salience of the distractor was a factor in the failure. We examined the influence of salience by replacing the salient distractors with less salient distractors. These were both less typical category exemplars than those used in Experiment 1 (e.g., “eclair” instead of “cake”) and contained less elaboration. The question is if readers will still fail to resolve the anaphor, as they did in Experiment 1, or if the salience of the distractor was critical.

Method

Participants. Forty-five undergraduates at the State University of New York at Binghamton participated in exchange for course credit. The data were replaced from four participants who made more than 25% probe or comprehension question errors and from two participants who did not follow instructions. Therefore the analyses were based on 39 participants.

Materials. There were 18 experimental passages, with three versions of each. A sample passage can be found in Table 4. The Distractor-Anaphor and Distractor-No Anaphor versions from Experiment 1 were used, except that the salient distractor was replaced with a less typical category exemplar that was mentioned only once. The third version was the No Distractor-Anaphor version from Experiment 2, in which the distractor information was replaced by filler information that was equated for length with the distractor elaboration but was unrelated to the anaphor. The last line of both the Distractor-Anaphor and No Distractor-Anaphor contained the categorical anaphor and the last line of the Distractor-No Anaphor version did not.
Rating experiment. To ensure that the distractors selected for this experiment were less typical than those used in Experiment 1, 27 undergraduates at the State University of New York at Binghamton were asked which of the two exemplars was more typical. One was the distractor from Experiment 1 and the other was the distractor selected for this experiment. For example, the question for the sample passage was “Which of the following is the more typical dessert? A) Eclair B) Cake.” The order of choices was counterbalanced across subjects. The distractors used in Experiment 1 were selected as more typical 86% of the time. Further, for every item, the Experiment 1 distractors were rated as more typical.

Design and procedure. The design and procedure were identical to those in Experiment 1.

Results and Discussion

Error rates and reading times. As in Experiment 2, there were no differences across versions in error rates ($F_s < 1$) or reading times, $F_1(2,76) = 1.23$, $MS_e = 31,443, p = .30$; $F_2(2,34) = .83$, $MS_e = 26,730, p = .45$.

Probe recognition times. The mean correct probe response times are presented in Table 2. These values were computed after discarding outliers (Tukey, 1977), 5.3% of the data. Most central to our question, probe reaction times were 52 ms shorter in the Distractor-Anaphor version than in the Distractor-No Anaphor version: 824 ms vs 876 ms. This difference was significant by subjects, $t_1(37) = 3.87$, $SEM = 13.36, p < .001$, and by items, $t_2(17) = 2.09$, $SEM = 30.88, p = .05$. In contrast to Experiment 1, readers were faster to respond yes to the probe word “tart,” for example, after reading “save room for the dessert” than after reading “congratulated John on the promotion.” This occurred despite the fact that the antecedent was substantially backgrounded and a same-category distractor intervened between the antecedent and the anaphor. However, unlike Experi-
ment 1, the distractor was not elaborated nor was it a highly typical member of the category. Although these data do not allow us to separate the effects of elaboration and typicality, in Experiment 1 their combined effect clearly served to divert activation from the antecedent to the distractor when the anaphor was processed.

Not only does the 52-ms facilitation effect contrast with the null effect in Experiment 1, but the facilitation of the probe following the Anaphor line was not significantly larger when there was no distractor present: The facilitation on the antecedent probe in the No Distractor-Anaphor version, compared with the Distractor-No Anaphor version, was 43 ms. Thus, not only can we conclude that the less salient distractor provided less interference than the more salient distractor, but the inclusion of the less salient distractor did not appear to have provided any interference at all. That is, having encountered a sentence describing Wanda’s trip to the bakery to buy eclairs did not hinder the accessibility of “tart” when readers encountered the anaphor, “the dessert.” Although Corbett (1984) did not include the same conditions, he also found an effect of typicality: There was no longer a slowdown on the anaphor when the antecedent was a highly typical example of the category and the distractor was atypical.

Because we did not separate the effects of elaboration from the typicality of the distractor, we can only conclude that the presence of a lengthy episode describing a highly typical same-category distractor was sufficient to prevent resolution. However, it is possible that all of these features were not necessary; resolution may have also been prevented if a subset of these factors were included. For example, consider an elaborated episode intervening between the antecedent and the anaphor that contained information from the same semantic field as the antecedent, but did not mention a particular exemplar. That is, the anaphor may have also not been resolved if the backgrounding simply described Wanda’s trip to the bakery to buy unspecified sweets for her daughter’s graduation party.

In summary, although this experiment does not allow us to specify the full set of factors that were responsible for the failure to resolve the anaphor in Experiment 1, the combined results of Experiments 2 and 3 do allow us to begin to establish some boundary conditions under which anaphoric inferences are drawn. Consistent with the many studies demonstrating that anaphoric inferences are drawn reliably, the conditions under which readers fail to resolve categorical anaphors appear to be limited. Not only was the resolution failure in Experiment 1 dependent on the inclusion of a same-category distractor, but it seems to have been dependent on the distractor being highly salient.

**EXPERIMENT 4**

In contrast to the assumption that anaphoric inferences are reliably drawn, the results of Experiment 1 established that readers sometimes leave categorical anaphors unresolved. In Experiments 4 and 5 we investigate the nature of the resolution failure.

In the current experiment, we ask if participants would have ultimately resolved the anaphor in Experiment 1, despite the inaccessibility of the antecedent, if they had more time. The anaphor was always the last word of the passage and the probe was presented only 500 ms after readers pressed the line advance key. Because it is not clear if readers wait until all processing is complete before pressing the advance key, the reduction in accessibility of the antecedent in the Distractor-Anaphor version in Experiment 1 may have simply slowed the resolution process rather than eliminated it. Thus, the antecedent may have begun to accrue activation as soon as the anaphor was encountered but the resolution process may not have been complete when the probe was presented. This may be because anaphor resolution is not complete until a word or two after the anaphor when resolution is difficult (Ehrlich & Rayner, 1983) or because of task demands associated with line-by-line reading and probe tasks. Concerns have recently been raised about the interpretation of a lack of facilitation of antecedent probes; Gordon and his colleagues have argued that these findings are the result of task-specific strategies (Gordon & Chan, 1995; Gordon, Hendrick, & Foster, 2000).
To address these concerns, in Experiment 4 reading time on a postanaphor line was used as the dependent measure instead of the recognition latencies used in Experiment 1. This both provides additional time for readers to complete the processing of the anaphor and eliminates concerns about task demands. Passages appeared in two versions (see Table 5): The Distractor-Anaphor version from Experiment 1 was used (now labeled the Distractor version), in which the antecedent was backgrounded by a highly salient distractor, as well as the No Distractor-Anaphor version from Experiment 2 (now labeled the No Distractor version), in which the distractor elaboration was replaced by neutral filler material. Both versions included the anaphor line. Given the results of Experiments 2, the anaphor should be resolved in the No Distractor version. This was used as a baseline.

In contrast to Experiment 1, there was no probe task and two lines were added to the end of the passages. A target line followed the anaphor line in both versions and explicitly mentioned the antecedent in the subject position. For example, after reading that Wanda said to save room for dessert, the target sentence stated “The tart looked perfect.” Reading time on this line was the dependent measure. If the inclusion of the salient distractor simply slowed the resolution process, but did not eliminate it, readers should have resolved the anaphor by the time they processed the target line; although it is possible that resolution is not completed on the anaphor itself, it is unlikely that it is delayed more than a word or two beyond the sentence boundary. Thus, reading times should be equal on the target line in the two versions. On the other hand, if the anaphor is not resolved in the Distractor version, even given the additional time and the elimination of the probe task, reading times should be longer on the target line; time should be needed to reinstate the concept “tart.” Finally, each passage ended with a post-
target line that was neutral with respect to the anaphor.

Method

Participants. Thirty-six undergraduates at the State University of New York at Binghamton participated as part of the requirement for an introductory psychology course. The data from three participants who made more than 25% comprehension question errors were replaced. Therefore data analyses were based on 33 participants.

Materials. There were 18 experimental passages, with two versions of each: Distractor and No Distractor. A sample passage can be found in Table 5. As in Experiment 1, the Distractor version contained an antecedent as well as an elaborated, highly typical distractor. Further, the passages contained the categorical anaphor. Two sentences were added to the end of the passage. The target line, which was the penultimate line, explicitly mentioned the antecedent. And the final, posttarget line was neutral with respect to the anaphor and was included simply to complete the passage. The No Distractor version was identical except that the description of the distractor was replaced by neutral filler material, which was the same number of lines as the distractor elaboration. Twenty-five filler passages were interspersed throughout the experimental passages.

Design. The design was the same as that in Experiment 2.

Procedure. The procedure was the same as in the previous experiments except that the probe task was omitted.

Results and Discussion

Mean reading times are presented in Table 6. These were calculated after discarding outliers, 6.7% of the data (Tukey, 1977). There were no reading time differences on the pretarget, anaphor line in the two versions; \( p = .7 \). In contrast, the pattern of reading times on the target line provides support for the conclusions from Experiment 1; the anaphor was not resolved when there was a salient distractor in the passage. Reading times on the target line were 99 ms longer in the Distractor version than in the No Distractor version, 1468 ms vs 1369 ms, respectively. This difference was reliable by subjects, \( t(32) = 4.11, SEM = 24.11, p < .001 \), and by items \( t_2(17) = 2.08, SEM = 42.38, p = .03 \). Following the anaphor, an explicit mention of the antecedent was read more slowly when a same-category distractor was included in the passage than when it was not. If the anaphor had been resolved in the Distractor version by the time participants read the target line, the antecedent should have been as active as it was in the No Distractor version, leading to similar reading times in the two versions. However, time was clearly needed to retrieve the antecedent concept from memory. Given that readers did not resolve the anaphor by the time they reached the second word of the following sentence, it is unlikely that they ever would have. In addition to ruling out the possibility that the reduction in the accessibility of the antecedent simply delayed the resolution process, the results of this experiment provide converging support for the finding of Experi-

<table>
<thead>
<tr>
<th>Version</th>
<th>Experiment 4</th>
<th>Experiment 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Anaphor line</td>
<td>Target line</td>
</tr>
<tr>
<td>Distractor</td>
<td>1697</td>
<td>1468</td>
</tr>
<tr>
<td>No Distractor</td>
<td>1718</td>
<td>1369</td>
</tr>
<tr>
<td>Difference</td>
<td>-21</td>
<td>99*</td>
</tr>
</tbody>
</table>

* \( p < .05 \).
Passages appeared in two versions (see Table 7), Distractor and No Distractor. The Distractor version was the same as that used in several of the previous experiments. In the sample passage, the tart was mentioned early in the passage and was followed by several lines of elaboration about the cake that Wanda needed for her daughter’s graduation party. This was backgrounded by a couple of lines and followed by the anaphor line. In contrast, in the No Distractor version, the high typical distractor (e.g., cake) was now used as the antecedent. That is, Wanda was making a cake for John’s promotion party instead of a tart. “Cake” was mentioned once (parallel to “tart” in the Distractor version) and was followed by neutral filler material which detailed Wanda’s efforts to buy her daughter a necklace for graduation. Because the antecedent was “cake” rather than “tart,” the target line (“The cake looked perfect”) was consistent with the passage. Therefore, if readers erroneously reinstated the distractor in the Distractor version, reading times should be equal on the target line in the two versions. In contrast, if readers did not mistakenly reinstate the distractor, reading times should be longer on the target line in the Distractor version because it is inconsistent with the passage.

Method

Participants. Thirty-three undergraduates at the State University of New York at Binghamton participated in exchange for course credit. The data of one participant who made more than 25% comprehension question errors were replaced. Therefore data analyses were based on 32 participants.

Materials. There were 18 experimental passages, with two versions of each: No Distractor and Distractor. A sample passage can be found in Table 7. In the Distractor version, the brief introduction was followed by the referent (e.g., tart), which was mentioned only one time. This was followed by an average of five lines of elaboration of the high typicality distractor (e.g., cake). In contrast, in the No Distractor version, after a brief introduction, the highly
typical category member (e.g., cake) was introduced as the referent. It was mentioned one time. In place of the distractor elaboration, this was followed by neutral filler material which was equated with the distractor elaboration in number of lines.

Both versions included a line that contained the categorial anaphor (“... save room for the dessert”) and a penultimate, target line that made reference to the high typicality category exemplar (“The cake looked perfect”). In the Distractor version, this was the distractor and as such constituted a contradiction; in the No Distractor version this was the referent and as such was consistent with the passage. Both versions ended with a neutral posttarget line. Twenty-five filler passages were interspersed among the experimental passages.

**Design and procedure.** The design and procedure were identical to those in Experiment 4.

---

**Results and Discussion**

Reading times are presented in Table 6. The first thing to note is that reading times did not differ on the anaphor line, $p > .5$. In contrast, reading times on the target line were 389 ms longer in the Distractor version than in the No Distractor version, 1900 ms vs 1511 ms, respectively. This difference was reliable by subjects, $t_1(31) = 6.84$, $SEM = 56.74$, $p < .001$, and by items, $t_2(17) = 4.17$, $SEM = 82.08$, $p < .001$. The slowdown in the Distractor version continued onto the posttarget line, 1995 ms vs 1770 ms. This difference was also reliable, $t_1(31) = 4.25$, $SEM = 53.06$, $p < .001$; $t_2(17) = 3.80$, $SEM = 75.03$, $p = .001$. The large slowdown in the Distractor version provides strong support for the conclusion that readers did not incorrectly instantiate the distractor when they read the anaphor. Although
“cake” must have received some activation when readers processed the anaphor, given the finding in Experiments 1 and 4 that a salient distractor interfered with the accessibility of the referent, readers clearly rejected the distractor as the referent. Given the slow reading times on the target line, they were clearly aware that it was inconsistent with the passage. If they had selected the distractor as the referent of the anaphor, reading time on this line should have been the same as in the No Distractor version.

Finally, comparing Experiments 4 and 5, it is interesting to note the large difference in effect sizes. In the current experiment, readers took almost 400 ms longer to read “The cake looked perfect” when the cake was not the referent of the anaphor as when it was. In contrast, in Experiment 4 readers took only about 100 ms longer to read “The tart looked delicious” when the anaphor had not been resolved than when it had. In the current experiment the large slowdown suggests that readers were unable to integrate the target line with the rest of the passage. In contrast, the smaller effect size in Experiment 4 suggests that although “tart” was not active when readers encountered the target line, they were able to retrieve the antecedent from memory and integrate the target line into the passage. Thus, although the contextual constraints reduced the accessibility of the referent so that the signal from the anaphor was not strong enough to retrieve it, the referent was not completely inaccessible; when readers encountered an explicit mention of the antecedent (e.g., the tart), with an additional 100 ms of processing time they were able to retrieve it from memory.

EXPERIMENT 6

The results of the first five experiments are inconsistent with the view that anaphors initiate a search of memory that is carried out to completion. Instead, it appears that when anaphoric inferences are not needed for comprehension, readers may not choose to invest additional resources to complete the inference.

In the final experiment, we examine a prediction implicit in this account: if readers had devoted additional resources, they would have resolved the anaphor. That is, as suggested in the conclusions of Experiment 5, although the contextual constraints reduced the accessibility of the antecedent, they did not make the antecedent impossible to retrieve. Rather, readers chose not to devote the additional resources. Presumably this resulted from the fact that passages were coherent even without making the connection between the anaphor and its referent. To test this prediction, an instructional manipulation was used to increase the amount of attention readers paid to the anaphor. The line that contained the anaphor was highlighted with asterisks and participants were told that some of the comprehension questions that followed the passages would involve information in the asterisked line. Across the experiment, 50% of the comprehension questions addressed information that was presented on an asterisked line, making it an efficient strategy to carefully encode that information. However, only 9 of 86 comprehension questions, or approximately 10%, specifically asked about an anaphor. In addition, there were numerous anaphors in the passages that were not queried in the questions. Thus, although the instructions encouraged participants to encode the anaphor line carefully, they did not go as far as asking readers to resolve the anaphors.

If readers resolve the anaphors, despite the same set of text factors that led to a resolution failure in Experiments 1 and 4, this would provide support for the conclusion that readers treated these inferences as optional and chose not to devote additional resources to reinstating the antecedents when their accessibility was reduced. In contrast, if the anaphors are still not resolved, this will suggest that the anaphor was an inadequate retrieval cue; even if readers had investigated additional resources to search for an appropriate antecedent, they might have been unsuccessful.

Method

Participants. Forty-eight undergraduates at the State University of New York at Binghamton participated in exchange for course credit. The data from 7 participants who made more than 25% probe or comprehension question er-
rors were replaced. Therefore data analyses were based on 41 participants.

_Materials._ There were 18 experimental passages, with two versions of each: Anaphor and No Anaphor. A sample passage can be found in Table 8. The materials were identical to the passages used in Distractor-Anaphor and Distractor-No Anaphor versions of Experiment 1, except that asterisks were placed around the last line of the passages. The probe word was always the antecedent concept. The filler passages were also identical to those used in Experiment 1 except that asterisks were placed around one randomly selected line in each passage. Two comprehension questions followed each passage. In contrast to the first five experiments, instead of having exclusively Yes–No comprehension questions, many of the questions were open-ended. In the experimental passages the first question was the target question and addressed information from the asterisked line; in the Anaphor version, it required readers to retrieve the antecedent. For example, in the sample passage, the target question was “What dessert did Wanda prepare for John?” in the Anaphor version. In the No Anaphor version, where the last line of the passage stated that Wanda “congratulated John on the promotion,” the target question was “What did Wanda do as everyone sat down to eat?” For the filler passages, the second of the two comprehension questions addressed information from the asterisked line.

_Design and procedure._ The design was identical to Experiment 1 except that there were two comprehension questions after each passage. The procedure was also the same as in Experiment 1 except that the cue “XXX” was on the screen for 750 ms before being replaced by the probe word. This was increased from the 500 ms used previously to avoid processing spill-over from the asterisked line to the probe task. Finally, instead of pressing a key to answer the comprehension questions, participants wrote their answers on a sheet of paper.

**Results and Discussion**

_Error rates._ Probe error rates were 13.6 and 14.4% in the Anaphor and the No Anaphor versions, respectively. These did not differ across versions, t’s < 1, and were similar to those in Experiment 1.

_Reading times._ Last line mean reading times are presented in Table 2. These were calculated

<table>
<thead>
<tr>
<th>TABLE 8</th>
</tr>
</thead>
</table>

Sample Text—Experiment 6

**Introduction and antecedent**

Wanda was throwing a surprise party for her best friend John. John had just been promoted to Vice President of the company and some of his close friends wanted to congratulate him. Wanda even made him a _tart_.

**Distractor elaboration**

She felt a little pressured because her daughter’s graduation was the next day and she needed to prepare for that occasion as well. She also still had to get decorations and stop at the bakery for a _cake_. Her daughter loved their _cakes_. They were filled with a very rich cream and had gooey dark chocolate frosting on them. Wanda decided to have the _cake_ decorated with white flowers and a white border.

**Wrap-up**

Wanda hoped John’s party would be fun. The guests arrived right on time. As everyone sat down to eat, Wanda

- Anaphor ending
  ****** said to save room for the _dessert_. ****
- No Anaphor ending
  ****** congratulated John on the promotion. ****

**Probe**

_TART_

**Comprehension question**

Was John recently fired?
after eliminating outliers (Tukey, 1977), 9.6% of the data. The first thing to note is that reading times were approximately 300 ms slower than in Experiment 1, providing evidence that readers were spending additional time on the asterisked lines, as they were instructed. In addition, reading times were quite variable on the target line; almost 10% of the data were classified as outliers. This was due to some participants spending as much as 4 or 5 s on the target line. Although this variability makes it difficult to draw strong conclusions from the reading time data, it is notable that readers spent 128 ms longer on the target line in the Anaphor version than in the No Anaphor version. This difference was reliable by subjects, $t_1(40) = 2.48$, $SEM = 51.57$, $p = .017$, but did not quite reach significance by items, $t_2(17) = 1.78$, $SEM = 70.96$, $p < .09$. This difference is in contrast with the results from Experiment 1 in which reading times were approximately equal in the two versions. Thus, not only did readers spend additional time processing the target line, as instructed, but this time was spent resolving the anaphor; otherwise reading times should have been about equal in the two versions.

**Probe recognition times.** Mean correct probe response times are presented in Table 2. These values were computed after discarding outliers (Tukey, 1977), 3.8% of the data. Although the passages were exactly the same as in Experiment 1, a different pattern of results emerged. In contrast to the null effect in Experiment 1, probe recognition times were faster in the Anaphor version than the No Anaphor version, 865 ms vs 907 ms, respectively. This 42-ms difference was significant by subjects, $t_1(40) = 2.70$, $SEM = 15.53$, $p = .01$, and marginally significant by items, $t_2(17) = 1.75$, $SEM = 23.32$, $p = .10$.

These data, as well as the last line reading time data, show that the failure to resolve the anaphor in Experiments 1 and 4 was due to the antecedent being difficult to access rather than to it being completely inaccessible. With the identical materials, participants reinstated the antecedent and resolved the anaphor when asked to pay attention to the anaphor line. This provides support for the conclusion that these inferences were not needed for maintaining the coherence of the passage, but were optional, or elaborative. Assuming that a reader’s goal is to build a coherent mental representation of the information described by the text, if the coherence of the passage depended on identifying the referent for an anaphor, readers should have devoted additional resources. And given the results of the current experiment, this would have been sufficient to resolve the anaphor. However, instead of devoting additional resources, readers simply left the anaphor unresolved and read on.

**GENERAL DISCUSSION**

Implicit in the study of anaphoric inferences has been the assumption that they are necessary inferences and are therefore reliably drawn. However, in a series of six experiments we examined the hypothesis that there are conditions under which anaphoric inferences are not drawn even when an antecedent is available. Further, we hypothesized that the probability of resolution should be a function of two factors: the extent to which resolution is necessary to create a coherent discourse representation and the accessibility of the antecedent. That is, if there are conditions under which readers’ standard of coherence is met without making the connection between an anaphor and its antecedent, it may be inefficient for readers to devote resources reinstating the antecedent. To test this hypothesis, we attempted to create passages in which noun phrases could either be interpreted referentially or without referring to an earlier mentioned concept and in which antecedents were difficult to retrieve.

Placing these questions within a bottom-up theory of discourse processing, such as Kintsch’s (1988) Construction-Integration model or Myers and O’Brien’s (1998) resonance hypothesis, we assume that anaphors act like any other input to memory. Thus, the same set of factors that influence the effectiveness of any memory retrieval cue should determine the accessibility of the antecedent: the match between the cue and target, the memory strength of the cue, and so on. Further, given this theoretical framework, after related concepts in memory are activated, in a subsequent stage
some of the reactivated concepts are integrated with the current text input. Although the activation process is initiated automatically, readers can continue processing, resampling memory with the current contents of working memory. We assume that continued processing is most likely to occur when a noun phrase cannot be easily integrated into the ongoing text representation.

The six experiments reported here provided support for our hypotheses. When readers encountered the anaphors in Experiments 1 and 4, the antecedent was not reactivated when a salient distractor intervened between the anaphor and its antecedent. In contrast, in Experiment 2, facilitation was found for the antecedent probe after the anaphor was read despite the fact that there was an extensive amount of text intervening between the anaphor and the antecedent. Facilitation was also found in Experiment 3, despite the fact that there was a less salient distractor intervening between the anaphor and the antecedent. Thus, the presence of the salient distractor in Experiments 1 and 4 clearly interfered with resolution; when the anaphor was read, its signal may have been divided between the referent and distractor, reducing the activation of the referent (Myers & O’Brien, 1998) by creating a fan effect (e.g., Anderson, 1976; Myers et al., 1984; Reder & Ross, 1983). As a result, the activation of the antecedent did not reach a critical threshold. Or, the activation of the antecedent may not have been sufficiently higher than that of the distractor. As Greene et al. (1992) suggested, failure to locate an antecedent may occur because no entity matches well or because several entities match about equally well.

Although we might expect referential distance to play a role in the accessibility of the antecedent, either because the strength of the memory traces decay with time (Kintsch, 1988) or because of interference from the intervening text (Myers & O’Brien, 1998), in Experiment 2 distance was not sufficient to prevent resolution. However, with a greater amount of intervening text or a stronger topic shift, the results might have been different. In addition, the same amount of referential distance might have been sufficient to disrupt resolution if the backgrounding material were more complex, if readers were attending to some other important detail of the passages, and so on. This points to what is undoubtedly a complex interaction between text factors, making it unsurprising that O’Brien et al. (1997) found evidence for non-resolution when there was substantial distance between anaphors and their referents, whereas we found that distance alone was not sufficient to prevent resolution. Similarly, Corbett (1984) concluded that the inclusion of a same-category distractor simply slowed the resolution process where we found that it prevented resolution. Although we examined the influence of a number of text factors in the current set of experiments, additional research is needed to fully understand the influence of factors such as typicality, referential distance, and elaboration, as well as their interaction, on the reactivation of backgrounded information.

Although these experiments began to narrow down the possibilities of what the resolution failure entailed, several possibilities remain. The results of Experiment 5 ruled out the possibility that readers actually resolved the anaphors by erroneously selecting the distractor. Thus, although the distractor must have received at least temporary activation, interfering with the accessibility of the antecedent, readers did not ultimately integrate it with the anaphor. This may have been the result of an integration stage, where “cake” was rejected as the dessert of the dinner party. Alternately, the cue to memory may have consisted of an entire proposition; rather than the cue being “dessert,” it may have been “the dessert at the dinner party.” Given this, although “cake” received some activation due to the partial match, this was not sufficient for full retrieval.

One possibility is that because the activation of the referent did not reach a critical threshold, readers treated the anaphoric noun phrase as a new entity. That is, although the noun phrase was used anaphorically by the character in the passage (i.e., Wanda knows that the dessert is a tart), it was not treated as an anaphor by readers. Although definite noun phrases often refer to entities that have already been established, it is
not uncommon for them to be used to refer to entities that are new (Fraurud, 1990; Poesio & Vieira, 1998). Thus, given the lack of activation of “tart,” readers may not have processed “the dessert” as an anaphor at all. They may have understood “the dessert” to be attributive, rather than referential, referring to some non-specific dessert.

Another possibility is that the referent was partially activated when the anaphor was encountered, as a result of the conceptual overlap between the anaphor and the referent. This partial retrieval may have caused readers to have a “feeling of knowing” (e.g., Hart, 1965) or a sense of familiarity. Although this subthreshold activation was not sufficient to retrieve the antecedent from long-term memory, it may have been sufficient to inform readers that an antecedent was present in the passage. Thus, readers may have been aware that “the dessert” was referring to a specific entity. But because it would have required additional resources to retrieve this entity, and because the passage was coherent without knowing what dessert Wanda was serving, readers simply moved on. This suggests that readers may process noun phrases as anaphoric, or referential, even without identifying a particular referent.

Finally, in this series of experiments we attempted to maximize the probability that passages would read smoothly even if readers did not connect the targeted anaphors with their intended antecedents. This was based on our hypothesis that when an antecedent is available in memory readers should only fail to resolve an anaphor when the antecedent is difficult to access and when its retrieval is not essential for comprehension. We selected anaphors that were constrained by the context and that could either be used referentially or to introduce a new concept. For example, because “dessert” is part of our dinner party schema, even if there had been no antecedent included in the passage, the definite noun phrase “the dessert” could have been used to refer to some nonspecific dessert. In contrast, if Wanda had said to leave room for “the frozen drinks,” an entity that is not part of our dinner party schema, readers might have devoted additional resources to draw a bridging inference or spent more time searching memory in an attempt to integrate this concept into the passage.

Although we did not systematically manipulate factors that should influence readers’ assessment of coherence in the same way that we manipulated factors that should influence the accessibility of the antecedent, the pattern of results suggests that the unresolved anaphors in Experiments 1 and 4 did not constitute a comprehension failure. Based on the lack of a slowdown in reading on the anaphor line, it does not appear that readers devoted additional resources in an attempt to resolve the anaphors. Assuming that readers’ goals are to construct a coherent representation of the text, it seems unlikely that they would have simply read on if they had found the narratives incoherent or anomalous. The results from Experiment 6 also provide support for this conclusion; antecedents were as difficult to access as they were in Experiments 1 and 4, but with additional processing readers successfully reinstated the antecedents. Thus, we conclude that readers did not devote these resources when reading the passages in Experiments 1 and 4. However, this discussion is speculative, and in the same way that we investigated the influence of a variety of factors on the accessibility of the antecedent, an investigation is needed of the influence of a variety of factors on readers’ assessment of coherence.

REFERENCES
Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The acti-


(Received July 13, 1999)

(Revision received February 18, 2000)